Fast decisions reflect biases; slow decisions do not

Samantha Linn [1,*] Sean D. Lawley,[1,†] Bhargav R. Karamched [2,3,4,‡] Zachary P. Kilpatrick [5,§] and Krešimir Josić [6,7,‖]

[1] *Department of Mathematics, University of Utah, Salt Lake City, Utah, USA*
[2] *Department of Mathematics, Florida State University, Tallahassee, Florida 32306, USA*
[3] *Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA*
[4] *Program in Neuroscience, Florida State University, Tallahassee, Florida 32306, USA*
[5] *Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado 80309, USA*
[6] *Department of Mathematics, University of Houston, Houston, Texas 77004, USA*
[7] *Department of Biology and Biochemistry, University of Houston, Houston, Texas 77004, USA*

(Received 2 January 2024; accepted 26 June 2024; published 12 August 2024)

Decisions are often made by heterogeneous groups of individuals, each with distinct initial biases and access to information of different quality. We show that in groups of independent agents who accumulate evidence the first to decide are those with the strongest initial biases. Their decisions align with their initial bias, regardless of the underlying truth. In contrast, agents who decide last make decisions as if they were initially unbiased and hence make better choices. We obtain asymptotic expressions in the large population limit quantifying how agents' initial inclinations shape early decisions. Our analysis shows how bias, information quality, and decision order interact in nontrivial ways to determine the reliability of decisions in a group.

DOI: 10.1103/PhysRevE.110.024305

I. INTRODUCTION

Evidence accumulation models describe how different organisms integrate information to make choices [1]. They capture the dynamics of decision making, including the trade-off between speed and accuracy [2–6]. Such models can also be used to understand how decisions are made in social groups, whether individuals observe each other's choices [7–10] or act independently [11].

Evidence accumulation is often modeled using biased Brownian motion with the quality of evidence determining the magnitude of drift and diffusion. A choice is triggered when the process crosses a threshold. This threshold controls the timing and accuracy of agent decisions, but questions remain about how the order of choices in a group is related to their accuracy [12]. Members of a group who access the highest-quality information will tend to make the fastest and most accuracte decisions [9]. However, even before accumulating evidence, humans and other animals often exhibit initial biases which may reflect previously acquired information but could also be erroneous. Biases could also be innate [13], due to mistaken. assumptions [14], or influenced by previous decisions [15]. Here we ask how individuals' initial biases in a group affect the order and accuracy of their choices. When is a decision driven by an agent's initial bias as opposed to accumulated evidence?

We show that in groups of agents who only differ in their initial biases, early decisions are made by agents with the most extreme predispositions. This effect occurs even for groups of modest size (e.g., $N = 10$) but intensifies in large groups and can be described with precise asymptotics. Early choices agree with the agents' initial bias, regardless of the evidence they can access. Early decisions are also noise driven: If the noisiness of the evidence accumulation process differs between agents, then the noisier agents often decide first, even when they start with a smaller bias. On the other hand, late decisions do not depend on the initial bias and are more likely to be correct. These effects hold generically but not in the case of initially unbiased agents [9].

II. MODEL DESCRIPTION

We first assume that each individual in a population of N agents decides between two hypotheses: H^+ or H^- by computing the conditional probabilities, $P(H^{\pm}|\text{evidence})$. In the limit of rapid, independent observations that provide weak evidence, the log likelihood ratio, or *belief*, of agent i in the group, $X_i = \log \frac{P(H^+|\text{evidence}_i)}{P(H^-|\text{evidence}_i)}$, evolves as biased Brownian motion [1,16] [see Fig. 1(a)],

$$dX_i = \mu_i\, dt + \sqrt{2D_i}\, dW_i, \tag{1}$$

where the drift, μ_i, and diffusion coefficient, D_i, capture the strength and noisiness of the evidence, respectively [17]. For all agents the correct choice ($H \in H^{\pm}$) is given by the sign of the drift ($\text{sgn}[\mu_i] = \pm 1$).

Although Eq. (1) can be derived as a model of an ideal decision maker, it is often used as a phenomenological model. The model can explain the variability in response times, the

*Contact author: samantha.linn@utah.edu
†Contact author: lawley@math.utah.edu
‡Contact author: bkaramched@fsu.edu
§Contact author: zpkilpat@colorado.edu
‖Contact author: kresimir.josic@gmail.com

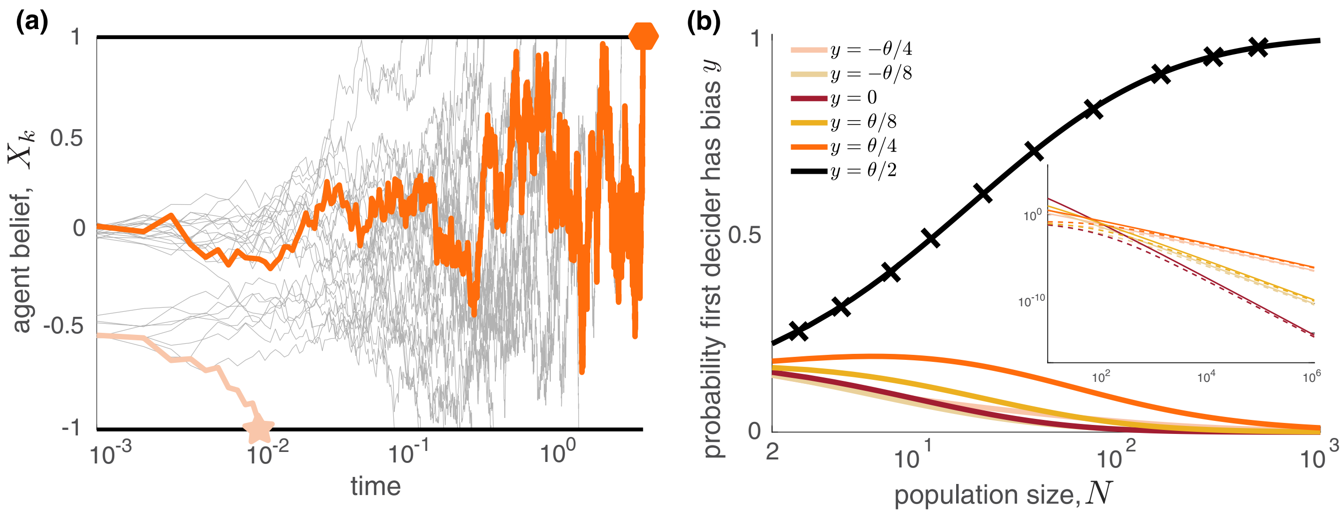

FIG. 1. Initial bias determines the choice of early deciders. (a) Evolution of a subset (20) of $N = 10^4$ agent beliefs, each of whom has even odds of initially being unbiased or biased [$P(X_j(0) = x_i) = 0.5$, $x_i = 0, -0.5$]. The first agent (light orange, star head) of $N = 10^4$ decides according to their initial bias and makes the wrong decision at $T_1 \approx 0.01$. The last agent (dark orange, hexagonal head) decides correctly at $T_{10^4} \approx 10$. (b) Probability that one of the agents with the largest initial bias decides first as a function of population size, N. Solid curves were determined by numerical quadrature [Eq. (A4) with initial biases $X_j(0)$ drawn with uniform probability from values listed in the legend]; black crosses denote results of a stochastic simulation with 10^6 trials. Inset: Log-log plot of the same results with asymptotics (dashed curves) computed from Eq. (3). Throughout, agents use identical thresholds $\pm\theta = \pm 1$, drift $\mu = 1$, and diffusivity $D = 1$.

speed-accuracy trade-off, as well as the impact of evidence quality and biases on choice [4,18,19]. Thus, initial bias, $X_i(0)$, cannot be assumed to be zero. The inherent biases of humans and other animals are difficult to eliminate or train away [13,20], reflect choice and reward history [15,19,21], reveal mistaken assumptions about the present choices [14], or arise from fluctuations or processes unrelated to the decision task.

We denote by y the initial data for a generic agent, while $X_i(0)$ is the initial bias of a specific agent. Agent i accumulates evidence, and its beliefs evolve according to Eq. (1), deciding when its belief reaches one of two thresholds, $\pm\theta$, at *decision time* $\tau_i := \inf\{t > 0 : X_i(t) \notin (-\theta, \theta)\}$. This decision, denoted by $d_i = H^{\pm}$, is determined by the sign of the threshold reached, $\mathrm{sgn}[X_i(\tau_i)]$. If decision criteria differ between agents, then an appropriate rescaling of $X_i(0)$, μ_i, and D_i allows us to assume that all agents use the same thresholds [1].

III. AGENTS WITH THE MOST EXTREME INITIAL BIASES DECIDE FIRST

We show that in large groups agents whose initial biases are closest to one of the thresholds make the earliest decisions. We first assume observers are identical except for their initial biases [$\mu_i = \mu$ and $D_i = D$ in Eq. (1)]. We denote by T_i the ith decision time so that $T_1 \leqslant T_2 \leqslant \cdots \leqslant T_N$, where $T_i = \tau_{n(i)}$ and $n(i)$ is the index of the ith agent to decide. Hence, the index of the first decider is $n(1)$.

For simplicity, we assume that each agent starts with one of finitely many initial beliefs, $\{x_0, x_1, \ldots, x_{I-1}\}$, sampled with probability $q_i = P(X_j(0) = x_i)$ for $i = 0, \ldots, I-1$. The distance of the initial belief x_i to the closest threshold is $L_i = \min\{\theta - x_i, x_i + \theta\}$. Let $i = 0$ be the index of the most extreme initial belief held by an agent so that $L_0 < L_i$ for $i \neq 0$.

For a fixed number of possible initial beliefs, I, the first to decide in a large group is an agent with the largest initial bias [Fig. 1(a)], in the sense that

$$P(X_{n(1)}(0) = x_0) \to 1 \quad \text{as } N \to \infty. \tag{2}$$

More precisely, in Appendix C we show that

$$P(X_{n(1)}(0) = x_i) \sim \eta_i (\ln N)^{(\beta_i - 1)/2} N^{1-\beta_i} \tag{3}$$

as $N \to \infty$ for each $i \neq 0$, where

$$\beta_i = (L_i/L_0)^2 > 1,$$
$$\eta_i = \frac{q_i}{q_0^{\beta_i}} \exp\left[\frac{\sqrt{\beta_i}}{2D}(\mu_i L_0 - \mu_0 L_i)\right]\sqrt{\beta_i \pi^{\beta_i - 1}}\,\Gamma(\beta_i) > 0,$$

and $\mu_i = \pm\mu$ if $x_i \gtrless 0$. The same statement holds if $n(1)$ is replaced by $n(j)$ in Eq. (3) but with a change in the prefactor, η_i (see Appendix C). Thus, the probability that the first decision is *not* made by the agent with the most extreme initial belief decreases as a negative power of the population size N [Fig. 1(b)]. The approximation given by Eq. (3) is in excellent agreement with the true probabilities when $N \gtrsim 10^3$ [see Fig. 1(b), inset]. Moreover, the probability that the agents with the most extreme initial beliefs make the first decision is close to unity already for $N \approx 10^2$ and occurs the majority of the time when $N \gtrsim 10$ as long as initial beliefs are well separated and drift is not too strong.

The choice of the fastest decider agrees with their initial bias, e.g., $P(X_{n(1)}(T_1) = \theta) \to 1$ as $N \to \infty$ if $\theta = \mathrm{argmin}_{x \in \pm\theta}|x_0 - x|$ [see Figs. 2(a) and 2(b)]. Similar results hold when initial beliefs are drawn from a continuous distribution (see Appendix E and next section): Early decisions of biased agents tend to be less accurate [11,22].

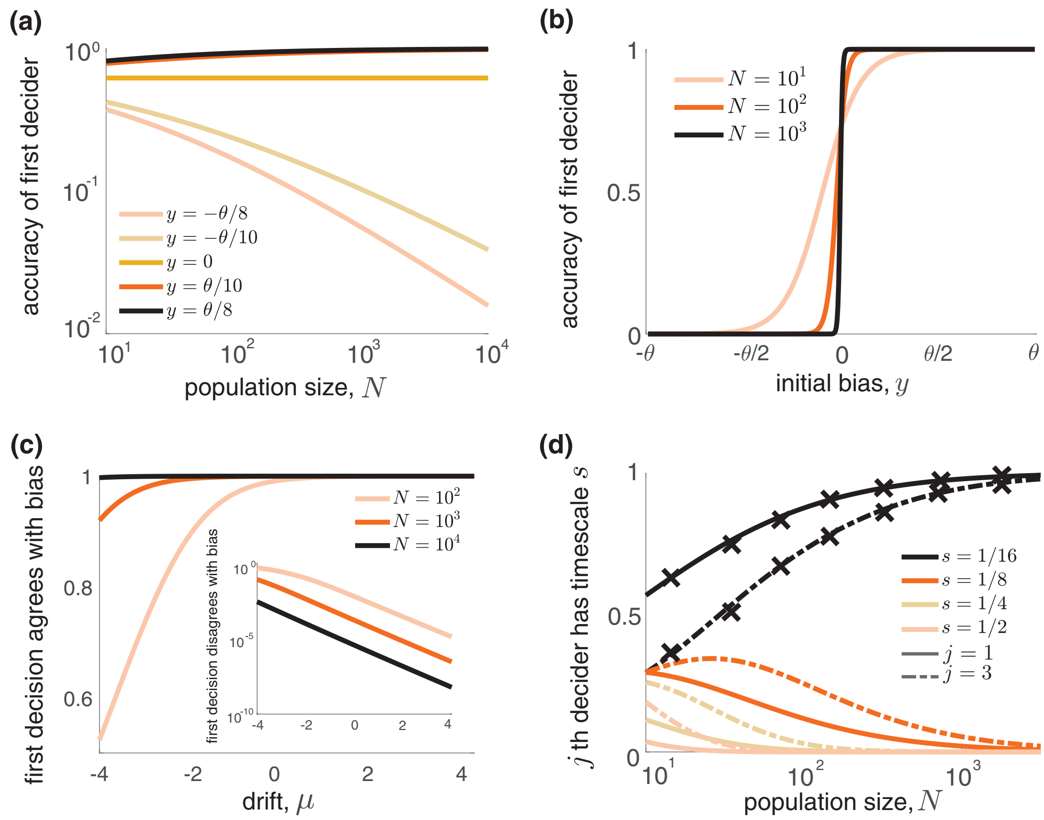

FIG. 2. First decider accuracy is determined by its initial bias. (a) The accuracy of the first decider as a function of population size, N, for different initial biases, y, obtained by quadrature. Curves are ordered by the proximity of the initial bias y of the first decider to the correct threshold $+\theta$. The drift, and hence the correct decision, are positive. Initial bias is uniformly distributed over legend values. (b) Under the same assumptions a small deviation from an unbiased initial belief strongly affects the probability of a correct first decision when N is large. (c) Drift weakly affects the first decision in populations with biased agents ($y = \theta/4$ here) when N is large. See Appendix D for decision polarity formulas. (d) In large populations in which all agents have the same initial bias, $y = \theta/2$, but different diffusivities, early deciders (here first and third) have the shortest diffusive timescale. X's mark averages of stochastic simulations over 10^6 trials.

In contrast, the probability that a single agent—or one chosen randomly without regard to decision order—decides incorrectly can be made arbitrarily small by increasing the drift or threshold [1]. In large populations with biased agents, drift and diffusion impact the probability of the first decision only through the prefactor in Eq. (3), η_i, and thus decrease in importance as population size diverges. If even a small part of a large population holds an initial bias, then early decisions are determined by the most extreme bias [Fig. 2(b)] regardless of the drift [Fig. 2(c)]. On the other hand, if all deciders are initially unbiased [$X_i(0) = 0$ for all i], then the probability that the first decider makes a correct choice is $[1 + \exp(-\mu\theta/D)]^{-1}$ [1].

IV. HETEROGENEOUS POPULATION AND CONTINUOUS DISTRIBUTION OF INITIAL BIASES

Our conclusions extend to populations of agents with heterogeneous distributions of initial biases, drifts, diffusivities, and thresholds. We again assume that each agent starts with one of finitely many initial beliefs, $X_i(0) \in \{x_0, x_1, \dots, x_{I-1}\}$ with drift and diffusivity sampled from a finite set of fixed size. For each agent we define the diffusive timescale,

$$S_i = \frac{L_i^2}{4D_i} > 0. \tag{4}$$

By assumption, the timescales S_i follow a discrete distribution with support on a finite set $0 < s_0 \leqslant s_1 \leqslant s_2 \leqslant s_3 \cdots \leqslant s_J$, and $S_{n(j)}$ refers to the timescale of the jth agent to decide [see Fig. 2(d)]. We denote by s the diffusive timescale of a generic agent.

In large populations, early deciders are those with the shortest diffusive timescales. In particular, we show in Appendix F that for every $\varepsilon > 0$ and *fixed* $j \geqslant 1$,

$$N^{1-s_1/s_0-\varepsilon} \ll P(S_{n(j)} > s_0) \ll N^{1-s_1/s_0+\varepsilon} \quad \text{as } N \to \infty, \tag{5}$$

where we use the notation $f \ll g$ to mean $\lim_{N\to\infty} f/g = 0$. We can thus conclude that $N^{1-s_1/s_0-\varepsilon} = o(P(S_{n(j)} > s_0))$ and $P(S_{n(j)} > s_0) = o(N^{1-s_1/s_0+\varepsilon})$ as $N \to \infty$.

These results agree with our earlier conclusion: If all agents share the same diffusivity, then the fastest deciders are the agents who start closest to their decision thresholds. This is true regardless of the quality of the evidence they receive. Diffusivity can reduce the effective distance to the threshold according to Eq. (4). The speed of the fastest deciders is

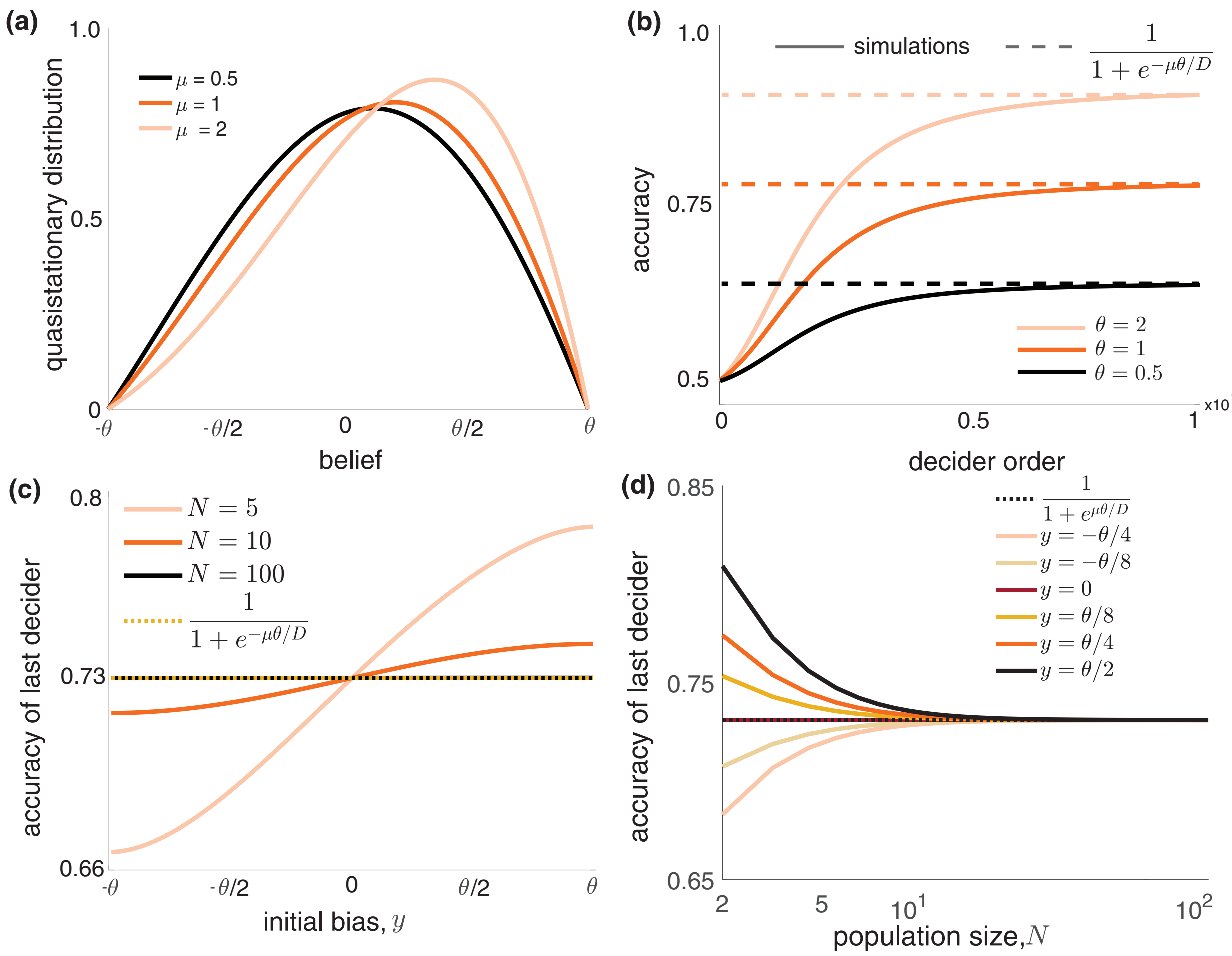

FIG. 3. Late deciders make choices as if they held no initial bias. (a) The quasistationary distribution $q(x)$ for various drifts μ represents the belief distribution for agents who remain undecided for a long time. (b) For large N, decision accuracy monotonically increases with decision order. The accuracy of late deciders approaches the accuracy of a single, initially unbiased agent. Here, all agents have initial bias $\theta/3$, and on each trial, $P(H = H^+) = 0.5$. (c) In large groups even a large initial bias does not impact the accuracy of later agents' choices. Here initial biases are sampled with uniform probability from $(-\theta, \theta)$. (d) The last decider begins to behave as if they were unbiased for $N = O(10)$. Unless otherwise noted, $\mu = 1$, $D = 1$, and initial beliefs are uniformly distributed over the values given in the legend.

determined by their initial proximity to threshold and noisiness of their integration process, regardless of the drift, μ_i. Our results also hold for models with nontrivial drift, like *leaky* Ornstein-Uhlenbeck models that describe integration in uncertain environments [23,24].

Our conclusions also pertain to agents with a continuous initial belief distribution. Suppose that an agent's initial belief is sampled from a smooth probability distribution, $\nu(x)$, with support on (a, b) with $-\theta < a < b < \theta$. Assuming

$$P(\tau \leqslant t | X(0) = x) \sim A(x) t^p e^{-C(x)/t}$$

as $t \to 0^+$ uniformly for all $x \in$ [a,b] where

$$C(x) = \frac{L(x)^2}{4D}, \quad L(x) = \min\{\theta - x, \theta + x\},$$

and $A(x) > 0$, it follows that (see Appendix E)

$$P(\tau \leqslant t) \sim \begin{cases} A(b)\nu_b \Gamma(\alpha_b + 1) t^{p+\alpha_b+1} e^{-C(b)/t} & \text{if } b > |a| \\ A(a)\nu_a \Gamma(\alpha_a + 1) t^{p+\alpha_a+1} e^{-C(a)/t} & \text{if } |a| > b, \end{cases}$$

where $\Gamma(z)$ denotes the gamma function. Define the event $E = \{a + \varepsilon < X(0) < b - \varepsilon\}$ $\quad 0 < \varepsilon \ll 1$ and suppose we want to estimate $\mathcal{P} \equiv P(a + \varepsilon < X_{n(1)}(0) < b - \varepsilon)$. We can show that (see Appendix E)

$$F_E(t) \equiv P(\tau \leqslant t \cap E)$$
$$\sim A(b - \varepsilon)\nu(b - \varepsilon) t^{p+1} e^{-C(b-\varepsilon)/t} \text{ as } t \to 0^+.$$

Substituting this into Theorem 1 in Appendix B shows that the first deciders have the most extreme beliefs. That is, $\mathcal{P} \to 0$ as $N \to \infty$.

V. LATE DECIDERS MAKE DECISIONS AS IF INITIALLY UNBIASED

We expect in large populations the inaccuracy of early deciders to be balanced by higher accuracy of late deciders [11]. Thus, we next determine the probability that the last agent to decide makes a correct decision. In Appendix G we show that this probability has an intuitive form,

$$P(X_{n(N)}(T_N) = \theta) \to \int_{-\theta}^{\theta} p_\theta(x)q(x)\,dx \quad \text{as } N \to \infty. \quad (6)$$

Here $p_\theta(x)$ is the probability that a single agent with initial bias $X(0) = x$ makes a correct decision, and $q(x)$ is the quasi-steady-state distribution [25] of beliefs evolving according to Eq. (1). Thus the decision of the last decider is made as if they forgot their actual initial bias and instead sample an initial belief from the quasistationary distribution, $q(x)$. Equation (6) is general and can be extended to arbitrary domains.

Figure 3(a) shows a examples of the quasi-steady-state distribution, $q(x)$. This is the distribution of beliefs of agents who have failed to make a decision for a long time and have "forgotten" their initial bias. Given this distribution, it is not immediately clear what the eventual decision of these late deciders will be as follows: Their beliefs, as captured by $q(x)$, favor the right decision, and the drift in Eq. (1) pushes these beliefs towards the correct threshold. Surprisingly, the resulting trajectories have the same probability of crossing the correct threshold as if the agents were initially unbiased, i.e., as if $P[X_{n(N)}](0) = 0$. More precisely, we show in Appendix G that $P(X_{n(N)}(T_N) = \theta) \to [1 + \exp(-\mu\theta/D)]^{-1}$ as $N \to \infty$, the probability that a single, initially unbiased decider makes a correct decision [see Figs. 3(b) and 3(d)] [1]. Thus, by forgetting their initial bias late deciders make decisions based only on accumulated evidence. The probability that an agent with a large initial bias makes a late decision is small. But should this happen, the initial bias will have little impact on their decision [see Fig. 3(c)].

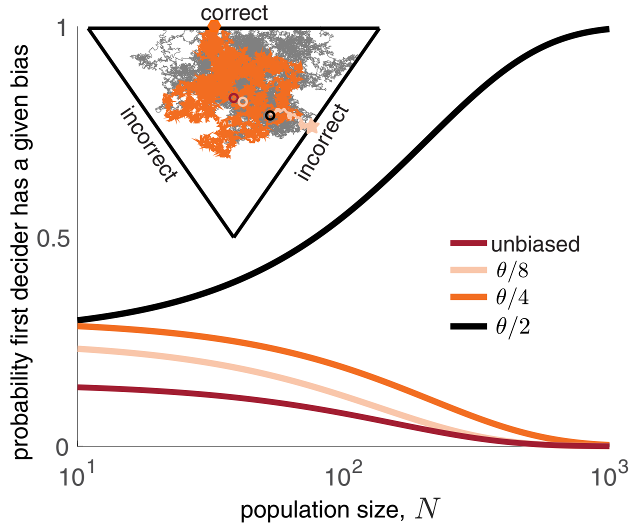

FIG. 4. Bias impacts multialternative and two-alternative decisions similarly in large groups. Beliefs about three options evolve on an equilateral triangle. Here θ is the closest distance from the center of the triangle (burgundy, central ring) to the boundary. The initial bias is the distance from the triangle center to the initial belief, $X_i(0)$. As N increases, the probability that the most biased agent chooses first grows. Curves are computed by averaging 10^6 stochastic simulations. Inset: Trajectories from a trial with biases sampled with equal probability from $\{\theta/2, \theta/4, \theta/8\}$. The first agent to decide (light orange, star head) has the largest initial bias. The belief of the last decider (dark orange, hexagonal head) explores the space before reaching a threshold.

VI. EXTENSION TO MULTIPLE ALTERNATIVES

We can extend these results to decisions between k alternatives. Equation (1) again describes the evolution of beliefs, but now $X_i(t)$, $\mu_i \in \mathbb{R}^{k-1}$ and W_i is a vector of independent Wiener processes [26]. Each belief evolves on a domain, $\Omega \subset \mathbb{R}^{k-1}$, with k boundaries [27], each associated with one of the alternatives. Agent i chooses alternative j if its belief, $X_i(t)$, crosses the associated boundary first. The boundaries that lead to the best decisions are difficult to find analytically [28], but their exact shape is immaterial for our result.

In Appendix F we show that Eq. (3) holds for general domains in arbitrary dimensions (see Fig. 4). We therefore reach our earlier conclusions: In large homogeneous populations, the agents holding the most extreme initial beliefs make the first decisions, and their choices are consistent with their initial biases. Our conclusions about the late decisions also carry over to agents facing multiple choices: Late deciders make choice as if they sampled their initial belief from the quasistationary distribution on Ω.

VII. DISCUSSION

Our decisions are often influenced by information we obtained previously and predilections we develop. In drift-diffusion models, prior evidence and predispositions can be represented by a shift in the initial state. We have shown that these initial biases determine early decisions and have diminishing influence on later decisions. Our main conclusions generally hold for populations of about a hundred, and our asymptotic results agree well with numerical solutions for populations larger than a few hundred agents. In extreme cases—when the drift, μ, is very large, or the distribution of initial beliefs very narrow—larger populations may be required for our observations to hold.

Agents in our model make decisions based on their initial beliefs, and accumulated evidence. The threshold that their belief crosses can be linked to the perceived accuracy of their decision [1], but the actual accuracy of their choice could be degraded by an erroneous initial bias and related to the temporal order of their decision. Though early decisions are not necessarily less accurate [7], our work identifies a clear case in which hasty choices tend to be the most unreliable. Our findings also suggest a means of weighting choices of biased agents by decision order in a large group to improve collective decisions [29].

In social groups the exchange of social information between agents [30,31] or correlations in the evidence [11] can affect these results. We have shown previously that knowing whether or not other agents have decided can be informative, even when no other social information is shared [22]. Here we assumed that each agent acts as if unaware of the others. It is well known that early decisions can trigger information cascades [32–34], and early adopters can have an outsized influence on the diffusion of technology and ideas in social networks [35]. Even weak social evidence can have a strong impact when shared early [7,36]. It remains unclear precisely how our results would be affected by such correlations, although we expect that under certain conditions they remain qualitatively true.

Our analysis of the fastest and slowest decisions joins several recent works which highlight the importance of extreme statistics in diverse biophysical systems. For example, the earliest receptor bindings may enable a single cell to locate a source [37,38] much more accurately than later receptor bindings [39]. The fastest receptor activations may also contribute to the effectiveness of kinetic proofreading for antigen discrimination by T cells [40], while the slowest primordial follicle growth activations determine menopause timing [41] and their extreme statistics shed light on the apparent "wasteful" follicle oversupply [42].

Ramping activity of individual neurons during decision making has been observed across the brain [43,44] (although see Ref. [45]). Such dynamics may reflect the underlying evidence accumulation process preceding a decision and is often modeled by a drift-diffusion process. Decisions are thought to be triggered by the elevated activity of sufficiently many choice-related neurons [46]. These results combined with our previous work on the impact of correlations [11] suggest that early decisions tend to exhibit lower accuracy. However, a key feature of neural circuits is their recurrent connectivity, which could help neural circuits reduce or even prevent the negative effects of extreme events [47].

Our theory also applies more generally to independently evolving drift-diffusion processes on possibly unbounded domains [48]: In large populations early threshold crossings reflect only the initial states, agnostic to other system attributes, while late crossings are independent of initial states and reflect the quasistationary distribution. Hence, early crossings reflect initial biases providing fast reactions needed for time-sensitive biophysical processes [49]. If time allows, then quorum sensing processes that weight passages by order could be used [50], managing population level trade-offs between speed and accuracy.

ACKNOWLEDGMENTS

Z.P.K. and K.J. were supported by NSF DMS-2207700. Z.P.K. was supported by NIH BRAIN 1R01EB029847-01. S.D.L. was supported by NSF DMS-2325258 and NSF DMS-1944574. K.J. was supported by NSF-DBI-1707400 and NIH RF1MH130416. S.L. was supported by NSF Grant No. 2139322.

APPENDIX A: MATHEMATICAL PRELIMINARIES

Suppose $\{(\tau_n, Z_n)\}_{n\geqslant 1}$ is an independent and identically distributed (iid) sequence of realizations of the pair of (possibly correlated) random variables (τ, Z). We have in mind that τ is the decision time [or first passage time (FPT)] of some decider whose stochastic evolution of beliefs is denoted by $\{X(t)\}_{t\geqslant 0}$ and Z is a vector containing information about this decider, such as their random initial position, drift, diffusivity, and decision made. Define the cumulative distribution function (CDF) of τ,

$$F(t) := P(\tau \leqslant t).$$

Further, for any event E that is in the σ algebra generated by Z, define

$$F_E(t) := P(\tau \leqslant t \cap E).$$

In words, E is any event for which we can know whether or not it occurred by knowing Z. For example, we are interested in events E like $E = \{X(0) = \theta/2\}$, $E = \{X(0) \leqslant 0\}$, $E = \{X(\tau) = \theta\}$, etc.

For a given $N \geqslant 1$, let $n(j) \in \{1, \dots, N\}$ denote the (random) index of the jth fastest decider out of the first N deciders to make a decision. That is, suppose we order the first N FPTs (or first decision times),

$$T_{1,N} \leqslant T_{2,N} \leqslant \cdots \leqslant T_{N-1,N} \leqslant T_{N,N},$$

where $T_{j,N}$ denotes the jth fastest FPT,

$$T_{j,N} := \min\left\{\{\tau_1, \dots, \tau_N\} \setminus \cup_{i=1}^{j-1} \{T_{i,N}\}\right\}, \quad j \in \{1, \dots, N\}. \tag{A1}$$

Then $n(j)$ is such that

$$\tau_{n(j)} = T_{j,N}. \tag{A2}$$

In the examples of interest, the FPTs, τ, have continuous probability distributions (i.e., $F(t)$ is a continuous function) so that the event $\tau_{n^*} = \tau_{n'} < \infty$ for $n^* \neq n'$ has probability zero so there is no ambiguity in Eq. (A2).

Since we have the sequence $\{(\tau_n, Z_n)\}_{n\geqslant 1}$, we denote as E_n the event E as it pertains to the nth element in the sequence $\{(\tau_n, Z_n)\}_{n\geqslant 1}$. For example, if $E = \{X(0) = \theta/2\}$, then $E_n = \{X_n(0) = \theta/2\}$. Similarly, $E_{n(j)}$ is the event E as it pertains to $Z_{n(j)}$.

Throughout the Appendix, we use the notation $\int f(t)\,dg(t)$ to denote the Riemann-Stieltjes integral of a function f with respect to a function g.

Proposition 1. For any $j \in \{1, 2, \dots, N\}$ (denoting an agent by the order j of their decision), we have that

$$P(E_{n(j)}) = j\binom{N}{j}\int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j}\,dF_E(t). \tag{A3}$$

In the case $j = 1$ (i.e., the fastest decider), Proposition 1 implies

$$P(E_{n(1)}) = N\int_0^\infty [1 - F(t)]^{N-1}\,dF_E(t). \tag{A4}$$

Since $1 - F$ is a decreasing function, Eq. (A4) implies that the short-time behavior of F and F_E determine the large N behavior of $P(E_{n(1)})$. More generally, Proposition 1 implies that the short-time behavior of F and F_E determine the large N behavior of $P(E_{n(j)})$ for $1 \leqslant j \ll N$.

In the case $j = N$ (i.e., the slowest decider), Proposition 1 implies

$$P(E_{n(N)}) = N\int_0^\infty [F(t)]^{N-1}\,dF_E(t). \tag{A5}$$

Since F is an increasing function, Eq. (A5) implies that the large-time behavior of F and F_E determine the large N behavior of $P(E_{n(N)})$. More generally, Proposition 1 implies that the large-time behavior of F and F_E determine the large N behavior of $P(E_{n(N-j)})$ for $1 \ll N - j$.

APPENDIX B: SOME INTEGRAL ASYMPTOTICS

The following proposition is useful for estimating the large-N behavior of some integrals of the form in Eq. (A3) and was proved in Ref. [48] (see Proposition 2 in Ref. [48]). Throughout the Appendix, "$f \sim g$" denotes $f/g \to 1$ (e.g., as $N \to \infty$ or as $t \to 0$).

Proposition 2. Assume $C_+ > C > 0$, $A > 0$, and $p, q \in \mathbb{R}$. Then there exists a $\delta_0 > 0$ so that for all $\delta \in (0, \delta_0]$, we have

$$\int_0^\delta t^{q-2} e^{-C_+/t} \big(1 - At^p e^{-C/t}\big)^{N-1}\, dt \sim \eta (\ln N)^{p\beta - q} N^{-\beta} \quad \text{as } N \to \infty,$$

where

$$\beta = C_+/C > 1, \quad \eta = C^{q-1}(AC^p)^{-\beta}\Gamma(\beta) > 0,$$

and $\Gamma(\beta) := \int_0^\infty z^{\beta-1} e^{-z}\, dz$ denotes the gamma function.

The following result estimates integrals of the form in Eq. (A3) for $1 \leqslant j \ll N$ assuming that $F(t)$ and $F_+(t)$ have short-time t behavior that is characteristic of diffusion.

Theorem 1. Assume $F(t)$ and $F_+(t)$ are bounded, nondecreasing, continuous from the right, and satisfy

$$F(t) \sim At^p e^{-C_0/t} \quad \text{as } t \to 0^+, \tag{B1}$$

$$F_+(t) \sim Bt^q e^{-C_+/t} \quad \text{as } t \to 0^+, \tag{B2}$$

where $C_+ > C_0 > 0$, $A > 0$, $B > 0$, and $p, q \in \mathbb{R}$. Then for any fixed integer $j \geqslant 1$, we have

$$j\binom{N}{j} \int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j}\, dF_+(t) \sim \eta(j)(\ln N)^{p\beta - q} N^{1-\beta} \quad \text{as } N \to \infty,$$

where

$$\beta := C_+/C_0 > 1, \quad \eta(j) := B(C_0)^{q-p\beta} A^{-\beta}\Gamma(j)\Gamma(\beta + j) > 0, \tag{B3}$$

and $\Gamma(x) := \int_0^\infty z^{x-1} e^{-z}\, dz$ denotes the gamma function.

Notice that the asymptotic behavior found in Theorem 1 as $N \to \infty$ is independent of $j \geqslant 1$, except for the constant prefactor $\eta(j)$. Further, this prefactor is an increasing function of j and satisfies

$$\eta(j) = \frac{(j-1)!\Gamma(\beta + j)}{\Gamma(\beta + 1)}\eta(1), \quad j \geqslant 1.$$

The asymptotic behavior in Eq. (B1) and (B2) is typical for diffusion, but computing the prefactors A and B and the powers p and q can be challenging [51]. Indeed, these constants depend on the details of the system (e.g., drift, space dimension, geometry of the domain, etc.). However, the constants in the exponents C_0 and C_+ are more universal and can be obtained in a very general mathematical setting [52]. The following result yields estimates on the fastest deciders when we only know these constants, which is equivalent to knowing the short-time behavior of $F_+(t)$ and $F(t)$ on a logarithmic scale.

Theorem 2. Assume $F(t)$ and $F_+(t)$ are bounded, nondecreasing, continuous from the right, and satisfy

$$\lim_{t \to 0^+} t \ln F(t) = -C_0 < 0, \quad \lim_{t \to 0^+} t \ln F_+(t) \leqslant -C_+ < 0, \tag{B4}$$

where $C_+ > C_0 > 0$. Then for every $\varepsilon > 0$,

$$j\binom{N}{j} \int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j}\, dF_+(t) = o(N^{1-\beta+\varepsilon}) \quad \text{as } N \to \infty, \tag{B5}$$

where

$$\beta := C_+/C_0 > 1.$$

If, in addition, we assume that

$$\lim_{t \to 0^+} t \ln F_+(t) = -C_+ < 0, \tag{B6}$$

then for every $\varepsilon > 0$,

$$N^{1-\beta-\varepsilon} = o\left\{ j\binom{N}{j} \int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j}\, dF_+(t) \right\} \quad \text{as } N \to \infty.$$

The following result estimates integrals of the form in Eq. (A3) for $1 \ll N - j \leqslant N$ assuming that $F(t)$ and $f_i(t) = F_i'(t)$ have large-time t behavior that is characteristic of diffusion in a bounded domain.

Theorem 3. Assume $F(t) \in [0, 1)$ is continuous and nondecreasing and $f_i(t)$ is continuous and bounded and

$$F(t) = 1 - ce^{-\lambda t} + \text{h.o.t.} \quad \text{as } t \to \infty, \qquad f_i(t) = \lambda c_i e^{-\lambda t} + \text{h.o.t.} \quad \text{as } t \to \infty,$$

where $\lambda > 0$, $c > 0$, $c_i > 0$. Then for any fixed $j \geqslant 0$, we have that

$$(N-j)\binom{N}{N-j}\int_0^\infty [F(t)]^{N-j-1}[1-F(t)]^j f_i(t)\,dt \to \frac{c_i}{c} \quad \text{as } N \to \infty.$$

APPENDIX C: PROOF OF EQ. (3)

We now apply Theorem 1 to obtain Eq. (3). Suppose the belief of each agent evolves independently according to the following stochastic differential equation (SDE),

$$dX = \mu\,dt + \sqrt{2D}\,dW, \tag{C1}$$

where $\mu \in \mathbb{R}$ is a constant drift, $D > 0$ is a constant diffusivity, and $W = \{W(t)\}_{t \geqslant 0}$ is a standard Brownian motion. Define the FPT,

$$\tau := \inf\{t > 0 : X(t) \notin (-\theta, \theta)\},$$

for some threshold $\theta > 0$. Assume that the initial distribution $P(X(0) = x_i)$ of each agent is a sum of Dirac masses at a finite set of points $\{x_0, x_1, \dots, x_{I-1}\}$,

$$P(X(0) = x) = \begin{cases} q_i & \text{if } x = x_i \text{ for some } i \in \{0, 1, \dots, I-1\}, \\ 0 & \text{if } x \notin \cup_{i=0}^{I-1} x_i. \end{cases},$$

Letting $F_i(t) \equiv F_{X(0)=x_i}(t) = P(\tau \leqslant t \cap X(0) = x_i)$, we have that [48]

$$F_i(t) \sim q_i A_i t^{1/2} e^{-C_i/t} \quad \text{as } t \to 0^+, \tag{C2}$$

where

$$C_i = \frac{(L_i)^2}{4D},$$

and

$$A_i = \begin{cases} \exp\left(\frac{-\mu L_i}{2D}\right)\sqrt{\frac{4D}{\pi (L_i)^2}} & \text{if } x_i < 0 \\ \exp\left(\frac{\mu L_i}{2D}\right)\sqrt{\frac{4D}{\pi (L_i)^2}} & \text{if } x_i > 0 \\ \left[\exp\left(\frac{-\mu L_i}{2D}\right) + \exp\left(\frac{\mu L_i}{2D}\right)\right]\sqrt{\frac{4D}{\pi (L_i)^2}} & \text{if } x_i = 0, \end{cases}$$

where L_i is the distance to the closest threshold from x_i,

$$L_i = \min\{\theta - x_i, x_i + \theta\}.$$

Further, we assume $0 \in \{0, 1, \dots, I-1\}$ is the index of the unique starting location closest to a threshold

$$L_0 = \min\{L_0, L_1, \dots, L_{I-1}\} < L_i \quad \text{if } i \neq 0,$$

then

$$F(t) \sim F_0(t) \quad \text{as } t \to 0^+.$$

We claim that

$$P(X_{n(1)}(0) = x_0) \to 1 \quad \text{as } N \to \infty, \tag{C3}$$

Thus, when N is large the first decider out of many deciders is always the one with the most extreme initial bias. Using the integral representation in Proposition 1 and applying Theorem 1 yields

$$P(X_{n(1)}(0) = x_i) \sim \eta_i(1)(\ln N)^{(\beta_i - 1)/2} N^{1-\beta_i} \quad \text{as } N \to \infty \quad \text{for each } i \neq 0,$$

where

$$\beta_i = (L_i/L_0)^2 > 1,$$

and

$$\eta_i(1) = \begin{cases} \frac{q_i}{q_0^{\beta_i}}\sqrt{\frac{\pi^{\beta_i-1}}{\beta_i}}\Gamma(\beta_i+1)\exp\left[\frac{\sqrt{\beta_i}}{2D}(\mu_i L_0 - \mu_0 L_i)\right] & \text{if } x_i \neq 0, \\ \frac{q_i}{q_0^{\beta_i}}\sqrt{\frac{\pi^{\beta_i-1}}{\beta_i}}\Gamma(\beta_i+1)\left\{\exp\left[\frac{\sqrt{\beta_i}}{2D}(\mu_i L_0 - \mu_0 L_i)\right] + \exp\left[\frac{\sqrt{\beta_i}}{2D}(-\mu_i L_0 - \mu_0 L_i)\right]\right\} & \text{if } x_i = 0, \end{cases}$$

where $\mu_i = \pm\mu$ if $x_i \gtrless 0$.

APPENDIX D: FIRST DECISION AGREES WITH INITIAL BIAS

The analysis above shows that the first agent to decide in a large group has the most extreme initial bias. We now show the intuitive result that this first decider's decision agrees with their initial bias. Without loss of generality, assume that the most extreme initial bias is negative, $x_0 < 0$. Letting $F_+(t) = P(\tau \leqslant t \cap X(\tau) = +\theta)$, we have

$$\begin{aligned} F_+(t) &= \sum_i P(\tau \leqslant t \cap X(\tau) = +\theta \,|\, X(0) = x_i)q_i \\ &\sim P(\tau \leqslant t \cap X(\tau) = +\theta \,|\, X(0) = x_{i^+})q_{i^+} \\ &\sim q_{i^+}A_{i^+}t_{i^+}^{p}e^{-C_{i^+}/t} \quad \text{as } t \to 0^+, \end{aligned}$$

where $i^+ \in \{1, \ldots, I\}$ is the index of the starting location closest to $+\theta$. Using the integral representation in Proposition 1 and applying Theorem 1 yields

$$P(X_{n(1)}(\tau) = +\theta) \sim \eta_{i^+}^{(1)}(\ln N)^{(\beta_{i^+}-1)/2}N^{1-\beta_{i^+}} \quad \text{as } N \to \infty.$$

APPENDIX E: CONTINUOUS INITIAL BELIEF DISTRIBUTION

In Appendix C, we showed that the first of many deciders have the most extreme initial beliefs in the case that the population has a discrete initial belief distribution. We now generalize this calculation to the case that the deciders have a continuous initial belief distribution. In particular, suppose that the decider's initial belief (position) has a smooth probability density $\nu(x)$ with support (a, b) with $-\theta < a < b < \theta$. Suppose that

$$\nu(x) \sim (x-a)^{\alpha_a}\nu_a \quad \text{as } x \to a^+, \qquad \nu(x) \sim (b-x)^{\alpha_b}\nu_b \quad \text{as } x \to b^-,$$

where the coefficients are positive, $\nu_a > 0$, $\nu_b > 0$, and the powers ensure that ν is integrable, $\alpha_a > -1$, $\alpha_b > -1$. In light of (C2), suppose that

$$P(\tau \leqslant t \,|\, X(0) = x) \sim A(x)t^p e^{-C(x)/t} \quad \text{as } t \to 0^+, \quad \text{uniformly for all } x \in [a, b],$$

where

$$C(x) = [L(x)]^2/(4D) > 0, \quad L(x) = \min\{\theta - x, \theta + x\},$$

and $A(x) > 0$ for all $x \in [a, b]$.

It follows that

$$F(t) = P(\tau \leqslant t) = \int_a^b P(\tau \leqslant t \,|\, X(0) = x)\nu(x)\,dx \sim t^p \int_a^b A(x)\nu(x)e^{-C(x)/t}\,dx \quad \text{as } t \to 0^+.$$

We thus need to estimate the small time t asymptotics of the integral

$$I := \int_a^b A(x)\nu(x)e^{-C(x)/t}\,dx,$$

which is an exercise in Laplace's method [53]. If $b > 0$, then for any $\varepsilon \in (0, b)$, we have

$$\int_0^b A(x)\nu(x)e^{-C(x)/t}\,dx \sim \int_{b-\varepsilon}^b A(x)\nu(x)e^{-C(x)/t}\,dx \sim A(b)e^{-C(b)/t}\nu_b\Gamma(\alpha_b+1)t^{\alpha_b+1} \quad \text{as } t \to 0^+.$$

Similarly, if $a < 0$, then for any $\varepsilon \in (0, |a|)$, we have

$$\int_a^0 A(x)\nu(x)e^{-C(x)/t}\,dx \sim \int_a^{a+\varepsilon} A(x)\nu(x)e^{-C(x)/t}\,dx \sim A(a)e^{-C(a)/t}\nu_a\Gamma(\alpha_a+1)t^{\alpha_a+1} \quad \text{as } t \to 0^+.$$

Putting this together, we have that if $b > |a|$, then

$$F(t) \sim A(b)\nu_b\Gamma(\alpha_b+1)t^{p+\alpha_b+1}e^{-C(b)/t} \quad \text{as } t \to 0^+,$$

and similarly if $|a| > b$ or $|a| = b$.

With these estimates, we can apply Theorem 1 to obtain estimates that the fastest decider(s) have extreme initial beliefs. In particular, suppose we want to estimate

$$P(a+\varepsilon < X_{n(1)}(0) < b-\varepsilon) \quad \text{for some small } 0 < \varepsilon \ll 1,$$

which is the probability that the fastest decider does not have extreme initial beliefs. If we define the event

$$E = \{a+\varepsilon < X(0) < b-\varepsilon\},$$

then using the notation of Appendix A, we have that

$$F_E(t) := P(\tau \leqslant t \cap E) = \int_{a+\varepsilon}^{b-\varepsilon} P(\tau \leqslant t \,|\, X(0) = x)\nu(x)\,dx \sim t^p \int_{a+\varepsilon}^{b-\varepsilon} A(x)\nu(x)e^{-C(x)/t}\,dx \quad \text{as } t \to 0^+,$$

which can be estimated as above using Laplace's method [53]. In particular, if $b > |a|$, then

$$F_E(t) \sim A(b-\varepsilon)\nu(b-\varepsilon)t^{p+1}e^{-C(b-\varepsilon)/t} \quad \text{as } t \to 0^+,$$

assuming $\nu(b-\varepsilon) > 0$, and similarly if $|a| > b$ or $|a| = b$. With this short-time behavior of $F_E(t)$, we can then plug this into Theorem 1 to show that the first deciders have the most extreme initial beliefs.

APPENDIX F: HETEROGENEOUS POPULATION WITH MULTIPLE ALTERNATIVES

We next consider the generalized case where the beliefs of the agents in the population evolve as processes with (possibly space-dependent) drift, diffusion coefficient, initial position, and even domain (in their own arbitrary space dimension $d \geqslant 1$). Suppose the belief of the ith decider evolves according to the following d-dimensional SDE,

$$dX_i = \mu_i(X_i)\,dt + \sqrt{2D_i}\,dW_i, \tag{F1}$$

where $\mu_i : \mathbb{R}^d \to \mathbb{R}^d$ is a possibly space-dependent drift, $D_i > 0$ is the diffusion coefficient, and $W(t) \in \mathbb{R}^d$ is a standard Brownian motion in d-dimensional space.

Let $L > 0$ denote an agent's (random) shortest distance they must travel to hit the closest target and let $D > 0$ denote the agent's diffusion coefficient. Define the random timescale

$$S = \frac{L^2}{4D} > 0.$$

Suppose that S has a discrete distribution on a finite set

$$0 < s_0 < s_1 < s_2 < s_3 \cdots < s_I,$$

where

$$P(S = s_i) = q_i > 0, \quad \sum_{i=0}^{I} q_i = 1.$$

Since we have $N \geqslant 1$ iid agents indexed from $n = 1$ to $n = N$, we let S_n denote the value of S for the nth agent and $S_{n(j)}$ the value of S for the jth fastest to decide.

We have that [52]

$$\lim_{t\to 0^+} t \ln P(\tau \leqslant t) = -s_0 < 0, \quad \lim_{t\to 0^+} t \ln P(\tau \leqslant t \cap S = s_i) = -s_i < 0.$$

Hence, Proposition 1 and Theorem 2 imply that for any fixed $j \geqslant 1$ and $i \in \{1, \ldots, I\}$ and any $\varepsilon > 0$,

$$N^{1-s_i/s_0-\varepsilon} \ll P(S_{n(j)} = s_i) \ll N^{1-s_i/s_0^{\varepsilon}} \quad \text{as } N \to \infty, \tag{F2}$$

where we use the notation $f \ll g$ to mean $\lim f/g = 0$. That is, in more traditional notation,

$$N^{1-s_i/s_0-\varepsilon} = o(P(S_{n(j)} = s_i)) \quad \text{as } N \to \infty,$$

$$P(S_{n(j)} = s_i) = o(N^{1-s_i/s_0+\varepsilon}) \quad \text{as } N \to \infty.$$

In the special case that the agents all move in one space dimension and the drifts are spatially constant (but may differ between agents), we can get the constant and logarithmic prefactors on the decay of $P(S_{n(j)} = s_i)$ as $N \to \infty$.

The result in Eq. (F2) says that in a large population if all the agents have the same diffusion coefficient, then the fastest deciders started closest to their decision thresholds (targets). If we allow the diffusion coefficients to vary between agents, then (F2) implies that the fastest deciders started close to their decision thresholds and/or they had big diffusion coefficients.

APPENDIX G: SLOWEST DECIDERS

Suppose the beliefs of the iid agents diffuse in some d-dimensional spatial domain $U \subset \mathbb{R}^d$ and can be absorbed at one of $m \geqslant 2$ targets $V_0, \ldots, V_{m-1}$ and let $\kappa \in \{0, \ldots, m-1\}$ indicate which target the decider eventually hits. Here, we will think of the m targets as parts of the $d-1$ dimensional boundary of the domain, and assume that hitting one of the targets triggers a decision. Following Refs. [54,55], suppose the beliefs of the deciders evolve as stochastic process $\{X(t)\}_{t\geqslant 0}$ that diffuse according to the SDE

$$dX(t) = -\nabla V[X(t)]\,dt + \sqrt{2D}\,dW(t), \tag{G1}$$

with reflecting boundary conditions. In Eq. (G1), the drift term is the gradient of a given potential, $V(x)$, and the noise term depends on the diffusion coefficient $D > 0$ and a standard d-dimensional Brownian motion (Wiener process) $\{W(t)\}_{t\geqslant 0}$. The survival probability conditioned on the initial position,

$$\mathbf{S}(x,t) := P(\tau > t \,|\, X(0) = x),$$

satisfies the backward Kolmogorov (also called backward Fokker-Planck) equation,

$$\begin{aligned}
\frac{\partial}{\partial t}\mathbf{S} &= \mathcal{L}\mathbf{S}, \quad x \in U,\\
\mathbf{S} &= 0, \quad x \in \text{targets},\\
\frac{\partial}{\partial \mathbf{n}}\mathbf{S} &= 0, \quad x \in \text{reflecting boundary (if there is one)},\\
S &= 1, \quad t = 0.
\end{aligned} \tag{G2}$$

In Eq. (G2), the differential operator $\mathcal{L}$ is the generator (i.e., the backward operator) of Eq. (G1),

$$\mathcal{L} = -\nabla V(x)\cdot\nabla + D\Delta,$$

and $\frac{\partial}{\partial \mathbf{n}}$ is the derivative with respect to the inward unit normal $\mathbf{n} : \partial U \to \mathbb{R}^d$.

Using the following weight function of Boltzmann form,

$$\rho(x) := \frac{e^{-V(x)/D}}{\int_U e^{-V(y)/D}\,dy}, \tag{G3}$$

one can check that the differential operator $\mathcal{L}$ is formally self-adjoint on the weighted space of square integrable functions (see, for example, Ref. [55]),

$$L^2_\rho(U) := \left\{ f : \int_U |f(x)|^2 \rho(x)\,dx < \infty \right\},$$

using the boundary conditions in (G2) and the following weighted inner product,

$$(f,g)_\rho := (f, g\rho) = \int_U f(x)g(x)\rho(x)\,dx,$$

where $(f,g) = \int_U f(x)g(x)\,dx$ denotes the standard L^2-inner product (i.e., with no weight function). Expanding the solution to (G2) yields,

$$\mathbf{S}(x,t) = \sum_{n\geqslant 1} (u_n, 1)_\rho e^{-\lambda_n t} u_n(x) = \sum_{n\geqslant 1} (u_n, \rho) e^{-\lambda_n t} u_n(x), \tag{G4}$$

where

$$0 < \lambda_1 < \lambda_2 \leqslant \ldots, \tag{G5}$$

denote the (necessarily positive) eigenvalues of $-\mathcal{L}$. The corresponding eigenfunctions $\{u_n(x)\}_{n\geqslant 1}$ satisfy the following time-independent equation:

$$-\mathcal{L}u_n = \lambda_n u_n, \quad x \in U, \tag{G6}$$

and identical boundary conditions as $\mathbf{S}$. Further, the eigenfunctions are orthogonal and are taken to be orthonormal, which means that

$$(u_n, u_m)_\rho = \delta_{nm} \in \{0, 1\}, \tag{G7}$$

where δ_{nm} denotes the Kronecker delta function (i.e., $\delta_{nn} = 1$ and $\delta_{mn} = 0$ if $n \neq m$).

If the initial distribution of an agent has probability measure μ_0,

$$P(X(0) \in B) = \mu_0(B) = \int_B 1\, d\mu_0(x), \quad B \subset U, \tag{G8}$$

then the FPT τ has survival probability given by

$$S(t) := P(\tau > t \,|\, X(0) =_d \mu_0) = \int_U \mathbf{S}(x,t)\, d\mu_0(x),$$

where the condition $X(0) =_d \mu_0$ in the conditional probability merely denotes that $X(0)$ has initial distribution given by μ_0. Hence, we obtain the following representation for the survival probability:

$$S(t) = \sum_{n\geqslant 1} A_n e^{-\lambda_n t} = \sum_{n\geqslant 1} (u_n, \rho)(u_n, d\mu_0) e^{-\lambda_n t}, \tag{G9}$$

where the coefficients are given by the following integrals:

$$A_n := (u_n, 1)_\rho \int_U u_n(x)\, d\mu_0(x), \quad n \geqslant 1. \tag{G10}$$

We have that the FPT τ to one of the targets has CDF

$$\begin{aligned} F(t) &= P(\tau \leqslant t) = 1 - P(\tau > t) \\ &= 1 - \sum_{k\geqslant 1} (u_k, \rho)(u_k, d\mu_0) e^{-\lambda_k t}, \end{aligned}$$

If

$$p_i(x) = P(\kappa = i | X(0) = x),$$

then

$$F_i(t) := P(\tau \leqslant t \cap \kappa = i) = P(\kappa = i) - P(\tau > t \cap \kappa = i) = P(\kappa = i) - \sum_{k\geqslant 1} (u_k, p_i\rho)(u_k, d\mu_0) e^{-\lambda_k t},$$

and therefore

$$f_i(t) := F_i'(t) = \sum_{k\geqslant 1} \lambda_k (u_k, p_i\rho)(u_k, d\mu_0) e^{-\lambda_k t}$$

Applying Proposition 1 and Theorem 3 yields

$$P(\kappa_{n(N-j)} = i) \to \frac{(u_1, p_i\rho)}{(u_1, \rho)} = \frac{(u_1\rho, p_i)}{(u_1, \rho)} \quad \text{as } N \to \infty.$$

Now, the solution to the forward Fokker-Planck equation is given by

$$p(x,t) = P(X(t) = dx \,|\, \tau > t) = \sum_{k\geqslant 1} e^{-\lambda_k t} (u_k, d\mu_0) \rho(x) u_k(x).$$

Hence, $u_1(x)\rho(x)/(u_1,\rho)$ is the quasistationary distribution (QSD), $q(x)$, defined by

$$\begin{aligned} q(x) &= \lim_{t\to\infty} P(X(t) = dx \,|\, \tau > t) = \lim_{t\to\infty} \frac{P(X(t) = dx \cap \tau > t)}{P(\tau > t)} = \lim_{t\to\infty} \frac{\sum_{k\geqslant 1} e^{-\lambda_k t}(u_k, d\mu_0)\rho(x)u_k(x)}{\sum_{k\geqslant 1}(u_k,\rho)(u_k, d\mu_0)e^{-\lambda_k t}} \\ &= \lim_{t\to\infty} \frac{e^{-\lambda_1 t}(u_1, d\mu_0)\rho(x)u_1(x)}{(u_1,\rho)(u_1, d\mu_0)e^{-\lambda_1 t}} = \frac{\rho(x)u_1(x)}{(u_1,\rho)}. \end{aligned}$$

Summarizing, we have shown that

$$P(\kappa_{n(N-j)} = i) \to \int_U p_i(x) q(x)\, dx \quad \text{as } N \to \infty. \tag{G11}$$

1. The case of drift-diffusion processes in one dimension

For the one-dimensional example in which all the beliefs of all the agents evolve according to (C1), we can compute the QSD, and find that

$$q(x) = \frac{\left(\pi^2 D^2 + \theta^2\mu^2\right)\cos\left(\frac{\pi x}{2\theta}\right) e^{\frac{\mu(\theta+x)}{2D}}}{2\pi D^2 \theta\left(e^{\frac{\theta\mu}{D}} + 1\right)}.$$

Further, it is straightforward to show that the probability that a decider reaches $+\theta$ before $-\theta$ conditioned on the initial belief $x \in [-\theta, \theta]$ is

$$p_1(x) := P(X(\tau) = +\theta) = \frac{1}{2}\left[\coth\left(\frac{\theta\mu}{D}\right) - 1\right] e^{\frac{\mu(\theta-x)}{D}}\left[e^{\frac{\mu(\theta+x)}{D}} - 1\right]$$

Therefore, applying (G11) and explicitly computing the integral yields

$$P(\kappa_{n(N-j)} = 1) \to \int_{-\theta}^{\theta} p_1(x)q(x)\,dx = \frac{1}{1+e^{-\frac{\theta\mu}{D}}} = p_1(0) \quad \text{as } N \to \infty.$$

Hence, the slowest deciders out of $N \gg 1$ deciders make a decision as if they were initially unbiased [i.e., as if $X(0) = 0$].

APPENDIX H: PROOFS

Proof of Proposition 1. Since $\{(\tau_n, Z_n)\}_{n\geqslant 1}$ are identically distributed, we have that

$$\begin{aligned} P(A_{n(j)}) &= \sum_{\substack{\text{distinct indices} \\ n_1,\dots,n_N \in \{1,\dots,N\}}} P(\max\{\tau_{n_1}, \dots, \tau_{n_{j-1}}\} < \tau_{n_j} < \min\{\tau_{n_{j+1}}, \dots, \tau_{n_N}\} \cap A_{n_j}) \\ &= j\binom{N}{j} P(\max\{\tau_1, \dots, \tau_{j-1}\} < \tau_j < \min\{\tau_{j+1}, \dots, \tau_N\} \cap A_j), \end{aligned} \tag{H1}$$

where the coefficient comes from noting that the number of terms in the sum is obtained by choosing the j fastest FPTs out of N and then choosing which of those j will be the jth fastest. Define

$$\tau_j^{(A_j)} = \begin{cases} \tau_j & \text{if } A_j \text{ occurs,} \\ +\infty & \text{if } A_j \text{ does not occur,} \end{cases}$$

so that if $j < N$,

$$P(\max\{\tau_1, \dots, \tau_{j-1}\} < \tau_j < \min\{\tau_{j+1}, \dots, \tau_N\} \cap A_j) = P(\max\{\tau_1, \dots, \tau_{j-1}\} < \tau_j^{(A_j)} < \min\{\tau_{j+1}, \dots, \tau_N\})$$

To handle the case $j = N$, we can simply replace $+\infty$ by $-\infty$ in the definition of $\tau_j^{(A_j)}$.

Since $\{\tau_n\}_{n\geqslant 1}$ are iid, we have that

$$P(\max\{\tau_1, \dots, \tau_{j-1}\} < t) = P(\max\{\tau_1, \dots, \tau_{j-1}\} \leqslant t) = [F(t)]^{j-1},$$

where we have used that $F(t)$ is continuous. Similarly,

$$P(\min\{\tau_{j+1}, \dots, \tau_N\} > t) = [1 - F(t)]^{N-j},$$

Using that $\{\tau_n\}_{n\geqslant 1}$ are independent, we have

$$\begin{aligned} G(t) &:= P(\max\{\tau_1, \dots, \tau_{j-1}\} < t < \min\{\tau_{j+1}, \dots, \tau_N\}) \\ &= P(\max\{\tau_1, \dots, \tau_{j-1}\} < t) P(t < \min\{\tau_{j+1}, \dots, \tau_N\}) \\ &= [F(t)]^{j-1}[1 - F(t)]^{N-j}. \end{aligned}$$

Combining the above finally yields

$$\begin{aligned} P(A_{n(j)}) &= j\binom{N}{j} P(\max\{\tau_1, \dots, \tau_{j-1}\} < \tau_j^{(A_j)} < \min\{\tau_{j+1}, \dots, \tau_N\}) \\ &= j\binom{N}{j} \mathbb{E}\big[G\big(\tau_j^{(A_j)}\big)\big] \\ &= j\binom{N}{j} \int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j}\,dF_E(t), \end{aligned}$$

which completes the proof.

The proof of Theorem 1 is similar to the proof of Theorem 3 in Ref. [48].

Proof. Define the integral from $t = a$ to $t = b$,

$$I_{a,b} := \int_a^b [F(t)]^{j-1}[1 - F(t)]^{N-j}\,dF_+(t).$$

Let $\varepsilon \in (0, 1)$. By the assumptions in Eq. (B1) and (B2), there exists a $\delta > 0$ so that

$$A_{-\varepsilon} t^p e^{-C_0/t} \leqslant F(t) \leqslant A_{+\varepsilon} t^p e^{-C_0/t} \quad \text{for all } t \in (0, \delta), \tag{H2}$$

$$B_{-\varepsilon} t^q e^{-C_+/t} \leqslant F_+(t) \leqslant B_{+\varepsilon} t^q e^{-C_+/t} \quad \text{for all } t \in (0, \delta), \tag{H3}$$

where $A_{\pm\varepsilon} := A(1 \pm \varepsilon)$ and $B_{\pm\varepsilon} := B(1 \pm \varepsilon)$. Using Eq. (H2) and integrating by parts yields

$$\begin{aligned}
I_{0,\delta} &\leqslant \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (1 - A_{-\varepsilon} t^p e^{-C_0/t})^{N-j} \, dF_+(t) \\
&= (A_{+\varepsilon} \delta^p e^{-C_0/\delta})^{j-1} (1 - A_{-\varepsilon} \delta^p e^{-C_0/\delta})^{N-j} F_+(\delta) \\
&\quad + (N-j) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (pt^{-1} + C_0 t^{-2}) A_{-\varepsilon} t^p e^{-C_0/t} \big(1 - A_{-\varepsilon} t^p e^{-C_0/t}\big)^{N-j-1} F_+(t) \, dt \\
&\quad - (j-1) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (pt^{-1} + C_0 t^{-2}) \big(1 - A_{-\varepsilon} t^p e^{-C_0/t}\big)^{N-j} F_+(t) \, dt \\
&\leqslant (A_{+\varepsilon} \delta^p e^{-C_0/\delta})^{j-1} (1 - A_{-\varepsilon} \delta^p e^{-C_0/\delta})^{N-j} F_+(\delta) \\
&\quad + (N-j) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j} (pt^{-1} + C_0 t^{-2}) \big(1 - A_{-\varepsilon} t^p e^{-C_0/t}\big)^{N-j-1} B_{+\varepsilon} t^q e^{-C_+/t} \, dt \\
&\quad - (j-1) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (pt^{-1} + C_0 t^{-2}) \big(1 - A_{-\varepsilon} t^p e^{-C_0/t}\big)^{N-j} B_{-\varepsilon} t^q e^{-C_+/t} \, dt,
\end{aligned} \tag{H4}$$

where we have used Eq. (H3) in the final inequality. The first term in the right-hand side of Eq. (H4) vanishes exponentially fast as $N \to \infty$. Using Proposition 2 to find the large N behavior of the second two terms in the right-hand side of Eq. (H4) and the fact that $I_{\delta,\infty}$ vanishes exponentially fast as $N \to \infty$ yields

$$\limsup_{N \to \infty} \frac{j \binom{N}{j} I_{0,\infty}}{\eta_j (\ln N)^{p\beta - q} N^{1-\beta}} \leqslant \frac{(1+\varepsilon)}{(1-\varepsilon)^\beta}.$$

The analogous argument yields the lower bound

$$\liminf_{N \to \infty} \frac{j \binom{N}{j} I_{0,\infty}}{\eta_j (\ln N)^{p\beta - q} N^{1-\beta}} \geqslant \frac{(1-\varepsilon)}{(1+\varepsilon)^\beta}.$$

Since $\varepsilon \in (0, 1)$ is arbitrary, the proof is complete.

Proof. Define the integral from $t = a$ to $t = b$,

$$I_{a,b} := \int_a^b [F(t)]^{j-1} [1 - F(t)]^{N-j} \, dF_+(t).$$

By Eq. (B4), there exists a $\delta > 0$ so that

$$e^{-(C_0+\varepsilon)/t} \leqslant F(t) \leqslant e^{-(C_0-\varepsilon)/t} \quad \text{for all } t \in (0, \delta), \tag{H5}$$

$$F_+(t) \leqslant e^{-(C_+-\varepsilon)/t} \quad \text{for all } t \in (0, \delta). \tag{H6}$$

Using Eq. (H5) and integrating by parts yields

$$\begin{aligned}
I_{0,\delta} &\leqslant \int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t} \big[1 - e^{-(C_0+\varepsilon)/t}\big]^{N-j} \, dF_+(t) \\
&= F_+(\delta) e^{-(j-1)(C_0-\varepsilon)/\delta} \big[1 - e^{-(C_0+\varepsilon)/\delta}\big]^{N-j} \\
&\quad + (N-j)(C_0+\varepsilon) \int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t} t^{-2} e^{-(C_0+\varepsilon)/t} \big[1 - e^{-(C_0+\varepsilon)/t}\big]^{N-j-1} F_+(t) \, dt \\
&\quad - \int_0^\delta (j-1)(C_0-\varepsilon) t^{-2} e^{-(j-1)(C_0-\varepsilon)/t} \big[1 - e^{-(C_0+\varepsilon)/t}\big]^{N-j} F_+(t) \, dt.
\end{aligned} \tag{H7}$$

The first term in the right-hand side of Eq. (H7) vanishes exponentially fast as $N \to \infty$. To handle the second term in the right-hand side of Eq. (H7), note that Eq. (H6) implies that

$$\begin{aligned}&\int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t} t^{-2} e^{-(C_0+\varepsilon)/t}\big[1-e^{-(C_0+\varepsilon)/t}\big]^{N-j-1} F_+(t)\,dt\\&\quad\leqslant \int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t} t^{-2} e^{-(C_0+\varepsilon)/t}\big[1-e^{-(C_0+\varepsilon)/t}\big]^{N-j-1} e^{-(C_+-\varepsilon)/t}\,dt.\end{aligned} \tag{H8}$$

Since the third term in the right-hand side of Eq. (H7) is nonpositive, applying Proposition 2 to Eq. (H8) and using Eq. (H7) and the fact that $I_{\delta,\infty}$ vanishes exponentially fast as $N \to \infty$ completes the proof of Eq. (B5).

If Eq. (B6) holds, then there exists a $\delta > 0$ so that

$$\begin{aligned}e^{-(C_0+\varepsilon)/t} &\leqslant F(t) \leqslant e^{-(C_0-\varepsilon)/t} \quad \text{for all } t \in (0,\delta),\\ e^{-(C_++\varepsilon)/t} &\leqslant F_+(t) \leqslant e^{-(C_+-\varepsilon)/t} \quad \text{for all } t \in (0,\delta).\end{aligned} \tag{H9}$$

Using Eq. (H9) and integrating by parts yields

$$\begin{aligned}I_{0,\delta} &\geqslant \int_0^\delta e^{-(j-1)(C_0+\varepsilon)/t}\big[1-e^{-(C_0-\varepsilon)/t}\big]^{N-j}\,dF_+(t)\\
&= F_+(\delta)e^{-(j-1)(C_0+\varepsilon)/\delta}\big[1-e^{-(C_0-\varepsilon)/\delta}\big]^{N-j}\\
&\quad+(N-j)(C_0-\varepsilon)\int_0^\delta e^{-(j-1)(C_0+\varepsilon)/t} t^{-2} e^{-(C_0-\varepsilon)/t}\big[1-e^{-(C_0-\varepsilon)/t}\big]^{N-j-1} F_+(t)\,dt\\
&\quad-\int_0^\delta (j-1)(C_0+\varepsilon) t^{-2} e^{-(j-1)(C_0+\varepsilon)/t}\big[1-e^{-(C_0-\varepsilon)/t}\big]^{N-j} F_+(t)\,dt\\
&\geqslant F_+(\delta)e^{-(j-1)(C_0+\varepsilon)/\delta}\big[1-e^{-(C_0-\varepsilon)/\delta}\big]^{N-j}\\
&\quad+(N-j)(C_0-\varepsilon)\int_0^\delta e^{-(j-1)(C_0+\varepsilon)/t} t^{-2} e^{-(C_0-\varepsilon)/t}\big[1-e^{-(C_0-\varepsilon)/t}\big]^{N-j-1} e^{-(C_++\varepsilon)/t}\,dt\\
&\quad-\int_0^\delta (j-1)(C_0+\varepsilon) t^{-2} e^{-(j-1)(C_0+\varepsilon)/t}\big[1-e^{-(C_0-\varepsilon)/t}\big]^{N-j} e^{-(C_+-\varepsilon)/t}\,dt.\end{aligned} \tag{H10}$$

The first term in the right-hand side of Eq. (H10) vanishes exponentially as $N \to \infty$. Using Proposition 2 to estimate the second two terms in the right-hand side of Eq. (H10) completes the proof.

Lemma 1. For fixed $j \in \{0, 1, \ldots\}$, $c > 0$, $\lambda > 0$, and $\delta > 0$, we have that

$$(N-j)\binom{N}{N-j}\int_{1/\delta}^\infty \big[1-ce^{-\lambda t}\big]^{N-j-1} e^{-(j+1)\lambda t}\,dt \to \frac{1}{\lambda c^{j+1}} \quad \text{as } N \to \infty.$$

Proof. Changing variables

$$u = 1-ce^{-\lambda t}, \quad du = \lambda c e^{-\lambda t}\,dt$$

yields

$$\begin{aligned}\int_{1/\varepsilon}^\infty \big[1-ce^{-\lambda t}\big]^{N-j-1} e^{-(j+1)\lambda t}\,dt &= \frac{1}{\lambda c^{j+1}}\int_{1-ce^{-\lambda t}}^1 u^{N-j-1}(1-u)^j\,du\\
&= \frac{1}{\lambda c^{j+1}}\left[\frac{(N-j-1)!j!}{N!} - \int_0^{1-ce^{-\lambda t}} u^{N-j-1}(1-u)^j\,du\right],\end{aligned} \tag{H11}$$

where we have used that $\int_0^1 u^{a-1}(1-u)^{b-1}\,du = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. Since the integral in Eq. (H11) vanishes exponentially fast, the proof is complete.

Proof. Let $\varepsilon \in (0, 1)$. By assumption, there exists $\delta > 0$ so that

$$\begin{aligned}1-(1+\varepsilon)ce^{-\lambda t} &\leqslant F(t) \leqslant 1-(1-\varepsilon)ce^{-\lambda t} \quad \text{for all } t \geqslant 1/\delta,\\ \lambda(1-\varepsilon)c_i e^{-\lambda t} &\leqslant f_i(t) \leqslant \lambda(1+\varepsilon)c_i e^{-\lambda t} \quad \text{for all } t \geqslant 1/\delta.\end{aligned}$$

Defining the integral from $t = a$ to $t = b$,

$$I_{a,b} := \int_a^b [F(t)]^{N-j-1}[1-F(t)]^j f_i(t)\,dt,$$

we therefore have that

$$(1-\varepsilon)^{j+1}\lambda c_i c^j \int_{1/\delta}^{\infty} [1-(1+\varepsilon)ce^{-\lambda t}]^{N-j-1} e^{-(j+1)\lambda t}\, dt \leqslant I_{1/\delta,\infty}$$
$$\leqslant (1+\varepsilon)^{j+1}\lambda c_i c^j \int_{1/\delta}^{\infty} [1-(1-\varepsilon)ce^{-\lambda t}]^{N-j-1} e^{-(j+1)\lambda t}\, dt.$$

Since $I_{0,1/\delta}$ vanishes exponentially fast as $N \to \infty$, Lemma 1 implies that

$$\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{j+1}\frac{c_i}{c} \leqslant \liminf_{N\to\infty}(N-j)\binom{N}{N-j}I_{0,\infty} \leqslant \limsup_{N\to\infty}(N-j)\binom{N}{N-j}I_{0,\infty} \leqslant \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{j+1}\frac{c_i}{c}.$$

Since $\varepsilon \in (0, 1)$ is arbitrary, the proof is complete.

APPENDIX I: NUMERICAL SOLUTIONS

Numerical solutions were computed via trapezoidal quadrature on Eq. (A3) in Proposition 1. In each set of dynamics, we rescaled the drift-diffusion process on $[-\theta, \theta]$ to the interval $[0, \ell]$. The probability density function for hitting the left boundary in this system is [56]

$$f_0(t) := \frac{d}{dt}F_0(t) = \exp\left(-\frac{\mu x_0}{2D} - \frac{\mu^2 t}{4D}\right)\frac{D}{\ell^2}\phi\left(\frac{Dt}{\ell^2}, \frac{x_0}{\ell}\right), \tag{I1}$$

where

$$\phi(s, w) := \begin{cases} \sum_{k=1}^{\infty} \exp(-k^2\pi^2 s)2k\pi \sin(k\pi w), \\ (4\pi s^3)^{-1/2}\sum_{k=-\infty}^{\infty}(w+2k)\exp\left[-\frac{(w+2k)^2}{4s}\right]. \end{cases} \tag{I2}$$

The expressions in Eq. (I2) are equivalent but have distinct utility: the top expansion converges quickly for large s while the bottom expansion converges quickly for small s. Hence, we utilize both expressions to more accurately compute probabilities associated with slow and fast deciders, respectively.

Integrating Eq. (I1) yields

$$F_0(t) = \int_0^t f_0(t')\, dt' = \exp\left(-\frac{\mu x_0}{2D}\right)\Phi\left(\frac{Dt}{\ell^2}, \frac{x_0}{\ell}\right)$$

with long- and short-time expansions of $\Phi(s, w)$ given by

$$\Phi(s, w) = \int_0^s \phi(s', w)\, ds'$$
$$= \begin{cases} \sum_{k=1}^{\infty}\{1-\exp[-(b+k^2\pi^2)s]\}\frac{2k\pi}{b+k^2\pi^2}\sin(k\pi w), \\ \sum_{k=-\infty}^{\infty}\frac{\mathrm{sgn}(2k+w)}{2}\left\{e^{-\sqrt{\frac{b}{D}}|2k+w|}\mathrm{erfc}\left(\frac{|2k+w|}{\sqrt{4s}}-\sqrt{bs}\right)+e^{\sqrt{\frac{b}{D}}|2k+w|}\mathrm{erfc}\left(\frac{|2k+w|}{\sqrt{4s}}+\sqrt{bs}\right)\right\}, \end{cases}$$

where $b = (\mu\ell/2D)^2$. By symmetry one can determine the corresponding probability density and cumulative distribution functions for hitting the right boundary. Altogether, we acquire long- and short-time expressions for the cumulative distribution function of an agent making a decision,

$$F(t) := F_0(t) + F_1(t),$$

where numerical solutions are illustrated, we use the short-time expressions of ϕ and Φ for $10^{-10} \leqslant t \leqslant 1$ and the complementary long-time expressions for $1 < t \leqslant 100$, discretizing each time interval into 10^3 log-spaced points. We consider 10^3 terms in each series expansion. Moreover, we take $\ell = 1$ and unless otherwise stated $D = 1$. Finally, where more than one but finitely many initial beliefs are considered, we scale the probability functions according to the corresponding initial distribution as outlined in Appendix C.

Specific details of figures with numerical solutions are as follows: In Fig. (b) we illustrate in color Eq. (A4) where $F_E = F$ as defined above with $X_{n(1)}(0) = y$. The black curve, which contains the remaining mass of the total probability, is computed as the sum of the colored curves subtracted from one. In Figs. 2(a)–2(c) we illustrate the probability that the first decider chooses the decision at $X(T_1) = \theta$ conditioned on having a particular initial bias. Hence, by definition of conditional probability, the numerical solutions are produced from quadrature on ratios of Eq. (A4) with $F_E = F_1$ in the numerator and $F_E = F$ in the denominator with $X_{n(1)} = y$. The inset of Fig. 2(c) is one minus the outset. In Fig. 2(d) we illustrate Eq. (A3) where $F_E = F$ and $S_{n(j)} = s$. In Fig. 3(b) we illustrate the probability that the last decider chooses the decision at $X(T_N) = \theta$ conditioned on having a particular initial bias. Similarly to Fig. 2(b), the numerical solutions are produced from quadrature on ratios of Eq. (A5) with

$F_E = F_1$ in the numerator and $F_E = F$ in the denominator with $X_{n(N)}(0) = y$.

APPENDIX J: AGENT-BASED STOCHASTIC SIMULATIONS

1. One-dimensional drift diffusion equation

To test the analytical solutions, we solved Eq. (1) in the main text using the Euler-Maruyama method, which describes the evidence accumulation process preceding binary decisions. In this approximation scheme, the true solution to the stochastic differential equation is approximated by a Markov chain Y constructed by setting $Y_0 = X(0)$ and updating Y according to the iterative scheme

$$Y_{n+1} = Y_n + \mu\Delta t + \sqrt{2D}\Delta W,$$

where $Y_n \equiv Y(n\Delta t)$ is the value of the Markov chain after the nth update, and the random variables ΔW are independent and identically distributed Gaussian random variables with mean 0 and variance Δt. The equations were integrated until $|Y_n|$ exceeded θ.

The temporal discretization, Δt, is user defined. As N grows, the time to first decision decays slowly. Thus, for large N, Δt must be taken to be sufficiently small for accurate representation of decision dynamics. For simulations here, we chose $\Delta t = 10^{-3}$ for $1 \leqslant N \leqslant 1000$. For $N > 1000$, we chose $\Delta t = N^{-1}$.

2. Two-dimensional drift diffusion equation

Decisions between three choices require a drift-diffusion model evolving on a planar domain [27]. Updating the discrete-time approximation of Eq. (F1) for each observer (dropping the i subscript) using Euler-Maruyama provides the following iterative scheme

$$Y_{n+1}^1 = Y_n^1 + \mu^1\Delta t + \sqrt{2D}\Delta W^1,$$
$$Y_{n+1}^2 = Y_n^2 + \mu^2\Delta t + \sqrt{2D}\Delta W^2,$$

where $Y_n^j = Y^j(n\Delta t)$ is the value of the belief after the nth update, the random variables ΔW^j are Gaussian random variables with mean 0 and variance Δt. Equations are integrated until the vector $(Y_n^1, Y_n^2)^T$ departs the triangular domain

$$\{(Y^1, Y^2) | Y^2 < h \,\&\, Y^2 > -2h(3Y^1 + 1) \,\&\, Y^2 > 2h(3Y^1 - 1)\},$$

where $h = (2\sqrt{3})^{-1}$. Choices of each agent are determined by whether the agent crosses the $Y^2 = h$ or $Y^2 = -2h(3Y^1 + 1)$ or $Y^2 = 2h(3Y^1 - 1)$ boundary. For simulations again we use $\Delta t = 10^{-3}$ for $1 \leqslant N \leqslant 1000$ and $\Delta t = N^{-1}$ for $N > 1000$.

For the 2D case in an equilateral triangle, the threshold θ is taken to be equal to the length of the apothem—defined as a line from the center of a regular polygon at right angles to any of its sides. Hence, an unbiased agent begins at the center of the equilateral triangle. We prescribe initial data for biased agents to be anywhere along an apothem except the center of the triangle.

[1] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks, Psychol. Rev. **113**, 700 (2006).

[2] L. Chittka, A. G. Dyer, F. Bock, and A. Dornhaus, Bees trade off foraging speed for accuracy, Nature (London) **424**, 388 (2003).

[3] W. T. Newsome, K. H. Britten, and J. A. Movshon, Neuronal correlates of a perceptual decision, Nature (London) **341**, 52 (1989).

[4] R. Ratcliff, A theory of memory retrieval, Psychol. Rev. **85**, 59 (1978).

[5] J. A. Swets, W. P. Tanner Jr, and T. G. Birdsall, Decision processes in perception, Psychol. Rev. **68**, 301 (1961).

[6] N. Uchida and Z. F. Mainen, Speed and accuracy of olfactory discrimination in the rat, Nat. Neurosci. **6**, 1224 (2003).

[7] B. Karamched, M. Stickler, W. Ott, B. Lindner, Z. P. Kilpatrick, and K. Josić, Heterogeneity improves speed and accuracy in social networks, Phys. Rev. Lett. **125**, 218302 (2020).

[8] R. P. Mann, Collective decision making by rational individuals, Proc. Natl. Acad. Sci. USA **115**, 10387 (2018).

[9] A. Reina, T. Bose, V. Srivastava, and J. A. R. Marshall, Asynchrony rescues statistically optimal group decisions from information cascades through emergent leaders, R. Soc. Open Sci. **10**, 230175 (2023).

[10] A. N. Tump, M. Wolf, P. Romanczuk, and R. H. J. M. Kurvers, Avoiding costly mistakes in groups: the evolution of error management in collective decision making, PLoS Comput. Biol. **18**, e1010442 (2022).

[11] M. Stickler, W. Ott, Z. P. Kilpatrick, K. Josić, and B. R. Karamched, Impact of correlated information on pioneering decisions, Phys. Rev. Res. **5**, 033020 (2023).

[12] B. Zandbelt, B. A. Purcell, T. J. Palmeri, G. D. Logan, and J. D. Schall, Response times from ensembles of accumulators, Proc. Natl. Acad. Sci. USA **111**, 2848 (2014).

[13] L. Busse, A. Ayaz, N. T. Dhruv, S. Katzner, A. B. Saleem, M. L. Schölvinck, A. D. Zaharia, and M. Carandini, The detection of visual contrast in the behaving mouse, J. Neurosci. **31**, 11351 (2011).

[14] J. M. Beck, W. J. Ma, X. Pitkow, P. E. Latham, and A. Pouget, Not noisy, just wrong: The role of suboptimal inference in behavioral variability, Neuron **74**, 30 (2012).

[15] A. E. Urai, J. W. De Gee, K. Tsetsos, and T. H. Donner, Choice history biases subsequent evidence accumulation, Elife **8**, e46331 (2019).

[16] J. I. Gold and M. N. Shadlen, Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward, Neuron **36**, 299 (2002).

[17] For an ideal observer, if only measurement noise is present, then drift and diffusion are equal in the continuum limit, $\mu_i = D_i = m_i$ [1,24]. We consider the more general case where internal noise can increase $D_i > \mu_i$.

[18] J. I. Gold and M. N. Shadlen, The neural basis of decision making, Annu. Rev. Neurosci. **30**, 535 (2007).
[19] M. J. Mulder, E.-J. Wagenmakers, R. Ratcliff, W. Boekel, and B. U. Forstmann, Bias in the brain: A diffusion model analysis of prior probability and potential payoff, J. Neurosci. **32**, 2335 (2012).
[20] I. Rentschler, M. Jüttner, A. Unzicker, and T. Landis, Innate and learned components of human visual preference, Curr. Biol. **9**, 665 (1999).
[21] S. Goldfarb, K. F. Wong-Lin, M. Schwemmer, N. E. Leonard, and P. Holmes, Can post-error dynamics explain sequential reaction time patterns? Front. Psychol. **3**, 213 (2012).
[22] B. Karamched, S. Stolarczyk, Z. P. Kilpatrick, and K. Josić, Bayesian evidence accumulation on social networks, SIAM J. Appl. Dyn. Syst. **19**, 1884 (2020).
[23] J. R. Busemeyer and J. T. Townsend, Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment, Psychol. Rev. **100**, 432 (1993).
[24] A. Veliz-Cuba, Z. P. Kilpatrick, and K. Josić, Stochastic models of evidence accumulation in changing environments, SIAM Rev. **58**, 264 (2016).
[25] This is the distribution of beliefs conditioned on the absence of a decision after a long time.
[26] We interpret beliefs as log-likelihood ratios. Therefore, with k alternatives beliefs evolve in $k-1$ dimensions.
[27] T. McMillen and P. Holmes, The dynamics of choice among multiple alternatives, J. Math. Psychol. **50**, 30 (2006).
[28] S. Tajima, J. Drugowitsch, N. Patel, and A. Pouget, Optimal policy for multi-alternative decisions, Nat. Neurosci. **22**, 1503 (2019).
[29] J. A. R. Marshall, G. Brown, and A. N. Radford, Individual confidence-weighting and group decision-making, Trends Ecol. Evol. **32**, 636 (2017).
[30] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, Bayesian learning in social networks, Rev. Eco. Stud. **78**, 1201 (2011).
[31] E. Mossel and O. Tamuz, Opinion exchange dynamics, Probab. Surv. **14**, 155 (2017).
[32] L. R. Anderson and C. A. Holt, Information cascades in the laboratory, Am. Eco. Rev. **87**, 847 (1997).
[33] A. N. Tump, T. J. Pleskac, and R. H. J. M. Kurvers, Wise or mad crowds? The cognitive mechanisms underlying information cascades, Sci. Adv. **6**, eabb0266 (2020).
[34] D. J. Watts, A simple model of global cascades on random networks, Proc. Natl. Acad. Sci. USA **99**, 5766 (2002).
[35] O. Dedehayir, R. J. Ortt, C. Riverola, and F. Miralles, Innovators and early adopters in the diffusion of innovations: A literature review, Int. J. Innov. Mgt. **21**, 1740010 (2017).
[36] A. V. Banerjee, A simple model of herd behavior, Q. J. Econ. **107**, 797 (1992).
[37] A. J. Bernoff, A. Jilkine, A. N. Hernández, and A. E. Lindsay, Single-cell directional sensing from just a few receptor binding events, Biophys. J. **122**, 3108 (2023).
[38] A. E. Lindsay, A. J. Bernoff, and A. N. Hernández, Short-time diffusive fluxes over membrane receptors yields the direction of a signalling source, R. Soc. Open Sci. **10**, 221619 (2023).
[39] S. D. Lawley, A. E. Lindsay, and C. E. Miles, Receptor organization determines the limits of single-cell source location detection, Phys. Rev. Lett. **125**, 018102 (2020).
[40] J. Morgan and A. E. Lindsay, Modulation of antigen discrimination by duration of immune contacts in a kinetic proofreading model of t cell activation with extreme statistics, PLoS Comput. Biol. **19**, e1011216 (2023).
[41] S. D. Lawley and J. Johnson, Slowest first passage times, redundancy, and menopause timing, J. Math. Biol. **86**, 90 (2023).
[42] S. D. Lawley and J. Johnson, Why is there an "oversupply" of human ovarian follicles? Biol. Reprod. **108**, 814 (2023).
[43] R. Kiani, T. D. Hanks, and M. N. Shadlen, Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment, J. Neurosci. **28**, 3017 (2008).
[44] M. N. Shadlen and W. T. Newsome, Motion perception: Seeing and deciding, Proc. Natl. Acad. Sci. USA **93**, 628 (1996).
[45] K. W. Latimer, J. L. Yates, M. L. R. Meister, A. C. Huk, and J. W. Pillow, Single-trial spike trains in parietal cortex reveal discrete steps during decision-making, Science **349**, 184 (2015).
[46] X.-J. Wang, Neural dynamics and circuit mechanisms of decision-making, Curr. Opin. Neurobiol. **22**, 1039 (2012).
[47] R. P. N. Rao, Bayesian computation in recurrent neural circuits, Neural Comput. **16**, 1 (2004).
[48] S. Linn and S. D. Lawley, Extreme hitting probabilities for diffusion, J. Phys. A: Math. Theor. **55**, 345002 (2022).
[49] D. S. Grebenkov, R. Metzler, and G. Oshanin, From single-particle stochastic kinetics to macroscopic reaction rates: Fastest first-passage time of n random walkers, New J. Phys. **22**, 103004 (2020).
[50] A. Koriat, When are two heads better than one and why? Science **336**, 360 (2012).
[51] S. D. Lawley, Distribution of extreme first passage times of diffusion, J. Math. Biol. **80**, 2301 (2020).
[52] S. D. Lawley, Universal formula for extreme first passage statistics of diffusion, Phys. Rev. E **101**, 012413 (2020).
[53] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory* (Springer Science & Business Media, New York, 2013).
[54] J. B. Madrid and S. D. Lawley, Competition between slow and fast regimes for extreme first passage times of diffusion, J. Phys. A: Math. Theor. **53**, 335002 (2020).
[55] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, Berlin, 2014), Vol. 60.
[56] D. J. Navarro and I. G. Fuss, Fast and accurate calculations for first-passage times in wiener diffusion models, J. Math. Psychol. **53**, 222 (2009).